ELSEVIER

Artificial Intelligence 101 (1998) 135–163

Artificial Intelligence

A belief network approach to optimization and parameter estimation: application to resource and environmental management

Olli Varis [1]

Laboratory of Water Resources, Helsinki University of Technology, P.O. Box 5300, FIN-02015 HUT, Helsinki, Finland

Received 11 January 1996; received in revised form 23 June 1997

Abstract

An approach to use Bayesian belief networks in optimization is presented, with an illustration on resource and environmental management. A belief network is constructed to work parallel to a deterministic model, and it is used to update conditional probabilities associated with different components of that model. The divergence between prior and posterior probability distributions at the model components is used as an indication on the inconsistency between model structure, parameter values, and other information used. An iteration scheme was developed to force prior and posterior distributions to become equal. This removes inconsistencies between different sources of information. The scheme can be used in different optimization tasks including parameter estimation and optimization between various policy options. Also multiobjective optimization is possible. The approach is illustrated with an example on cost-effective management of river water quality. © 1998 Elsevier Science B.V. All rights reserved.

Keywords: Bayesian methods; Belief networks; Environmental policies; Hybrid models; Parameter estimation; Probabilistic models; Optimization; Resource management; Water quality

1. Introduction

Uncertainty is among the most discussed topics in environmental and resource management. Interest in probabilistic assessment, risk analysis, and related techniques has grown

[1] Email: ovaris@leka.hut.fi.

0004-3702/98/$19.00 © 1998 Elsevier Science B.V. All rights reserved.
PII: S0004-3702(98)00010-1

rapidly in the recent years [1,5,45]. Probabilistic and risk analyses are increasingly accepted in practical assessment work by international organizations and by national authorities in many countries. Risk-conscious, typically risk-averse approaches such as the precautionary principle have been endorsed by numerous governments. Modern decision theory, together with various recently developed computational techniques for processing uncertain information, provide a wide base for novel, potential approaches to applications in the field. At present, these opportunities are far from being properly known and fully utilized.

The concept of uncertainty has several facets in this context. From a decision-theoretic view, uncertainty can be grouped in three clusters [8,43]: (1) acquisition, presentation, and propagation of information available; (2) preferences and objectives of a given problem; and (3) structural issues. Pearl [19] divides computational techniques into two groups: logic-based approaches (monotonous logic in rule-based systems, etc.) and probabilistic ones (Bayes, Dempster–Shafer, fuzzy set theory, etc.). In this study, Bayesian calculus is used because it is known to have a strong theoretical basis and to provide an unified approach to statistical and deterministic theories, and to questions of testing and estimation [9].

Within environmental and resource management, the applications of Bayesian analysis have been dominated by classical Bayesian inference, i.e., parameter estimation, in which the Bayesian analysis is restricted to the parameter space. In decision theory, the idea of considering the entire model as a construct subject to uncertainty and subjectivity stem from the game theory of the 1930s and 1940s [25]. Games evolved into sequential games against uncontrolled 'nature', and abstractions such as decision trees were developed. Bayesian decision theory gradually gained increasing notice and emphasis [44]. These theoretical concepts were not developed into more applicable ones until the late 1960s [8,16,20].

Further development has been linked with advances in related computational mathematics [2,21,26]. Artificial intelligence has had a rapidly growing impact within the last ten years. A set of probabilistic, Bayesian-type approaches applicable or potentially applicable to decision analysis under high uncertainty has emerged [7,19,25,30]. Characteristic of these techniques—known as belief networks, causal networks, Bayesian nets, qualitative Markov networks, influence diagrams, or constraint networks—is the network presentation of interdependencies between probabilistic variables. The local-updating principle used allows construction of large and densely coupled networks in a practically realizable way and to operate interactively and on-line. In recent years, they have spread quickly to many application areas, including fault diagnosis, reliability theory, medicine, and pattern recognition.

According to Bobrow [3], a particularly successful technique has been the belief network approach by Pearl [17,19], which was also used in the present study. Szolovits and Pauker [30] stated that "... *Pearl's formulation has had a revolutionary impact on much of AI*". As is usual in such techniques, the entire model—*the hypothesis space*—is subjected to Bayesian analysis, not only the parameter space (cf. [6,18,27]). In contrast to classical probability theory, different sets of outcomes are allowed for related nodes, yielding an evident violation of the Kolmogorov axiomatization of the Bayes formula, yet Pearl [17] strongly argues against this very axiomatization: "*It is not hard to see that this textbook*

view of probability theory presents a rather distorted view of human reasoning and misses its most interesting aspects." Many decision analytic approaches have also been in line with these ideas (see, for instance, [4,22,23]).

Varis [33] examined Pearl's methodology [17–19], and offered suggestions for making the approach more suitable for decision analysis in resource management and environmental studies. The suggested approach has adapted ideas particularly from Bayesian decision analysis and from some common practices within the field. The most essential suggestion was that nodes can be linked in two layers: (1) the probabilities of all outcomes (all possible state values) of the each model variable (node) can be propagated using belief-function calculus; and (2) the outcomes can be linked using deterministic equations (algebraic or logical). This implies that the network is understood as an approximate, numerical approach to updating the uncertainty in different parts of the model, making probabilistic simulations (such as Monte Carlo analysis) useless. This updating works instantly and does not require off-line simulation runs.

In the present approach, the basic uncertainty propagation scheme is from Pearl. Yet, several extensions were developed to it, in order to provide a wider practical applicability:

- *Direction specific link*: instead of $\boldsymbol{M}_{i|j} = \boldsymbol{M}_{j|i}$ that Pearl uses (e.g., pp. 158–159 and continuation in [19]), also $\boldsymbol{M}_{i|j} \neq \boldsymbol{M}_{j|i}$ is allowed.
- *The link strength approach*: one parameter can be used to define the link matrix.
- *Negative links*: the interconnection between two variables can also be negative.
- *Node dependency level*: the sum of the link strength parameters of links to a node define how dependent the node is from the rest of the network.

The goal of this study was to formulate and test the use of this approach in optimization and parameter estimation. The basic concept—stemming from the idea of using a belief network parallel to a deterministic model to handle uncertainties in different parts of the model—is to look at inconsistencies between the *model* outcomes and *external information* such as management targets (cost levels, environmental indices, etc.) or observations to which the model should be fitted. Inconsistencies are shown by diverging prior and posterior probabilities in control variables (such as parameters or, say, wastewater treatment levels). In addition, certain properties of the links are adjusted empirically. An iteration scheme was developed for this purpose.

All three categories of uncertainty listed above (i.e., propagation and presentation, objectives and preferences, and structure) are supported. Uncertain information is propagated using discrete belief-function calculus. As far as the presentation and analysis of uncertainties in objectives and preferences are concerned, the discrete probabilistic domain allows the use of many concepts of utility theory, including risk-attitude analysis and value-of-information analysis. Structural uncertainty is handled in the two-layered model in the following manner: first, the variables can be linked by deterministic equations, and second by a network of conditional probabilities. This structure allows a degree of belief to be assigned to a deterministic dependency between variables. The approach also provides a number of possibilities to use models from different modeling traditions [34,36,39] within one meta-model.

The approach can also be understood as a generalized, discrete Kalman filter, in which also the state equation uncertainty—the structural uncertainty of the deterministic model—

is estimated. Bayesian influence diagrams [22,23] have been previously been used in filtering approaches [46], but without the structural uncertainty property, and without the bi-directional propagation scheme.

The approach is illustrated by a series of numerical examples, and a case study of water policy analysis of a river basin. This work is in continuum with a number of policy studies on resource and environmental management using and developing Bayesian decision analysis techniques and knowledge engineering approaches. The methodology has included influence diagrams [12,37,38,41], belief networks [13,34,35,39,40,42], and probabilistic risk analysis [10,31]. The applications have been on river basin management, lake water quality management, observational design, global change including climatic change studies, and fisheries management.

2. The uncertainty balance approach

Assume that we must solve a complex control and/or diagnosis problem with high uncertainties. The available information comes from diverse sources and is contradictory. A balanced view of the problem based on all the information sources is needed.

2.1. *Structure, targets, and uncertainty balance*

We have or we construct *a model* to describe the crucial elements of the problem (the term model has a slightly different meaning for a statistician and a deterministic modeler, but I try to fuse and merge these two concepts, and try to include both of these meanings in the term model). The model could be, for instance, a set of differential equations that describe the dynamics of organic pollution and dissolved oxygen in a river, or in more general, a relatively simple management oriented tool. We want to use this structure as a basis for our reasoning. In addition, we have *information that is external to the model* (knowledge, experience, data, goals, etc.). All the information is uncertain. We want to put this diverse, uncertain, and contradictory information into an analytic framework in which a reasonable compromise and balance between different pieces of information can be found.

There are observations and goals (= targets), with other possible external information that can be used together with the model. Technically, the approach divides the model into two layers that communicate with one another. The deterministic equations constitute *the state layer*, since it includes the state equations. It could also be called *the outcome layer*, because from that layer, one can get numerical values for the model variables (e.g., oxygen levels in a river). The other, *probabilistic layer*, consists of a network of approximate, conditional probability distributions for the outcomes.

2.2. *Noninformative network implies full balance*

How to use belief networks to assist in parameter estimation and, more generally, when optimizing control variables to fulfill the targets defined? The key proposition is that the prior and posterior probability distributions of the target variables (observations,

management goals, constraints, etc.) should become equal. This implies that the joint distributions of the external information (presented as prior distributions) should be equal to the modeled distributions of these variables (appearing as the posterior probability distributions), and assures that the prior information is properly utilized. This is done by value iteration in which the priors converge stepwise to the optimal values. This is in analogy to any error minimization procedure; here the error is shown as a difference between priors and posteriors.

The belief network is constructed so that if there is no external information diverging from the model prediction, then all the discrete probability distributions in the network are noninformative, uniform distributions. If the probabilistic layer consists of uniform distributions only, this tells us that no information is available other than that provided by the model, or, if there is external information, it agrees fully with the model.

2.3. *New information induces a need to re-establish the balance using control variables and links*

Introduce now a new piece of external, probabilistic information on targets in the analysis. Its probability distribution is approximated with a discrete distribution in which the outcomes are the same as in the corresponding model component, but in general the probability values become different.

The probabilistic layer is used to propagate this new information throughout the model. Evidently, all distributions deviating from the uniform distribution indicate that the model and the external information do not match completely. A controversy exists and it needs to be analyzed, and a proper balance should be found.

This can be done by adjusting the decision/control variables. They can be, e.g., parameters used to fit the model to data (= match targets), or wastewater treatment plants along a river to be upgraded to various purification levels to improve (= control) water quality in the river (a target again). In the latter, another set of targets may be the costs involved, and a balanced situation between these, typically contradictory targets should be found.

According to the proposition made in the previous section, the balance can be found by forcing the distributions calculated by the probabilistic layer (= posterior distributions) to be uniform. This implies that the joint distributions of the external information are equal to that of the modeled information. This can be achieved by changing the probabilities for outcomes of the control variables under consideration, until this goal is attained. The form of the posterior distribution gives a clear indication of how these distributions can be found. Another set of components that can be controlled to achieve the balance are the parameters describing how strongly two variables are interlinked. If, for instance, a link strength (terms including this one will be defined mathematically in the next section) corresponding to a deterministic model equation $= 1$, then we assume that this equation is 100% adequate in describing the phenomenon it should describe. If the link strength $= 0$, then we assume that the equation tells nothing on the phenomenon. Moreover, these link strengths clearly influence the model uncertainty calculated at the state layer. The lower the link strength is, the further the error bounds are from the expected behavior of the system. The reason

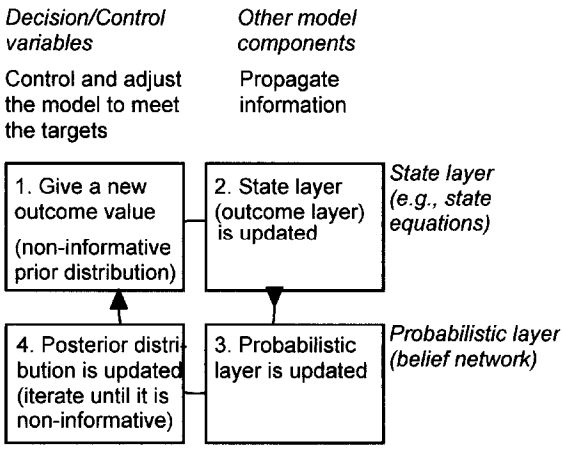

Fig. 1. Outline of the uncertainty balance iteration.

for this is that link strengths enable us to take into account the structural uncertainty of the model.

2.4. *Iteration for balance*

We want to achieve a situation in which the joint probability distributions of all probabilistic, external information propagated into control variables equal the prior distributions of these variables. The search is done by *uncertainty balance iteration* (Fig. 1). An intrinsic component of the analysis is the analyst her/himself, because much of the benefit of such analyses in nontrivial problems comes from the learning from and interaction with the information available. Therefore, the approach has been designed to be as interactive as possible and to be operated on-line.

3. Computational solution

3.1. *Propagation in state layer (outcome layer; deterministic state equations)*

If the deterministic state equations are nonlinear, as very often in practice, the analytical propagation of uncertainty is usually too laborious. Yet, there are many approximate approaches that can be used [11,15]. One of the most widely used ones is the Taylor series expansion. The more accuracy is required, the more terms can be included. We consider here the first-order approximation, which in many cases is sufficiently accurate. For equations expressing the deviations in output y from its nominal value, caused by deviations of $x_1, \ldots, x_n$ from their nominal values, the first-order approximation for the variance of y is

$$\mathrm{var}[y] \approx \sum_{i=1}^{n} \mathrm{var}[x_i] \left[\frac{\partial y}{\partial x_i} \right]^2. \tag{1}$$

There are two specific cases in which rather practical equations for expected value and uncertainty of y can be derived: the weighted sums of components and products of powers of the components. In the case of weighted sums

$$y = \sum_{i=1}^{n} a_i x_i \tag{2}$$

the mean and the variance can be obtained by

$$E[y] = \sum_{i=1}^{n} a_i E[x_i], \tag{3}$$

$$\mathrm{var}[y] = \sum_{i=1}^{n} a_i^2 \mathrm{var}[x_i] + 2 \sum_{i=1}^{n} \sum_{j=i+1}^{n} a_i a_j \mathrm{cov}[x_i, x_j]. \tag{4}$$

Accordingly, for product and power equations

$$y = \prod_{i=1}^{n} x_i^{a_i} \tag{5}$$

the mean and the variance are

$$E[y] \approx \prod_{i=1}^{n} E[x_i]^{a_i}, \tag{6}$$

$$\mathrm{var}[y] \approx \sum_{i=1}^{n} \mathrm{var}[x_i] \left(\frac{a_i}{x_i^0} y^0 \right)^2. \tag{7}$$

The variance equation can be processed in a more convenient form by using the coefficient of variation (cv)

$$\mathrm{cv}(z)^2 \equiv \frac{\mathrm{var}[z]}{E^2[z]}$$

hence,

$$\mathrm{cv}^2[y] \approx \sum_{i=1}^{n} a_i^2 \mathrm{cv}^2[x_i]. \tag{8}$$

Above, it was assumed that the model is structurally correct. In the present approach this does not need to be the case. As will be shown later, an uncertainty estimate can be given of the model structure, expressed as link strength η, a parameter defining a link matrix in the probabilistic layer. Details are given in Section 3.3. The link strength can be augmented to the state layer in the following approximate manner:

$$\mathrm{cv}'[y] \approx \frac{\mathrm{cv}[y]}{\sqrt{\eta}}, \tag{9}$$

where $\mathrm{cv}'[y]$ is the coefficient of variation of the model prediction when structural uncertainty is included. In $\mathrm{cv}[y]$ it is excluded.

3.2. Information from state layer to probabilistic layer

The normally distributed model prediction (Fig. 2) is approximated by a discrete distribution with n equally likely intervals. Hence, in a network with no measured information, all distributions are uniform. If any external information (measurement, target level, etc.) differing from the model prediction is included, then nonuniform distributions reveal it in the net. A further, practical rationale for using uniform distributions in the probabilistic layer is that a vector product of two discrete uniform distributions is a uniform distribution. This feature is important when propagating information in the probabilistic layer, as shown later.

Since the state layer uses continuous distributions and the probabilistic layer is in discrete form, discrete approximations of the continuous random variables are needed when taking them as priors to the probabilistic layer. The following approximation is used. First, define y_1 and y_2 such that

$$\begin{aligned} &P(y \leqslant y_1) = 1/3, \\ &P(y_1 < y \leqslant y_2) = 1/3, \\ &P(y > y_2) = 1/3. \end{aligned} \tag{10}$$

These values can be obtained by, e.g., using standard normal deviates:

$$\begin{aligned} y_1 &= \mu_y - 0.4307\sigma_y, \\ y_2 &= \mu_y + 0.4307\sigma_y. \end{aligned} \tag{11}$$

In other words, the model prediction is approximated here with a discrete distribution with three equally likely intervals. Also other number could be used. These values can then be used to find the discrete approximation of the evidence vector $\boldsymbol{e}$. This will now be made using the intervals obtained above.

$$\begin{aligned} &P(e_1) = P(e \leqslant y_1), \\ &P(e_2) = P(y_1 < e \leqslant y_2), \\ &P(e_3) = P(e > y_2). \end{aligned} \tag{12}$$

These values are used as the evidence vector in the probabilistic layer.

3.3. Propagation in probabilistic layer (belief network)

The belief network approach used in the propagation of information in the probabilistic layer is based on Pearl's work [19] with a set of extensions (see Introduction). In precise, Eqs. (14)–(21) are adopted from Pearl.

The probabilistic layer (belief network) consists of nodes connected with links.

Nodes. Each node i in a network contains:

- A vector of possible (discrete) outcomes (state values) $\boldsymbol{y}_i$ that can be defined as inputs, or they may depend on the outcomes of other nodes.

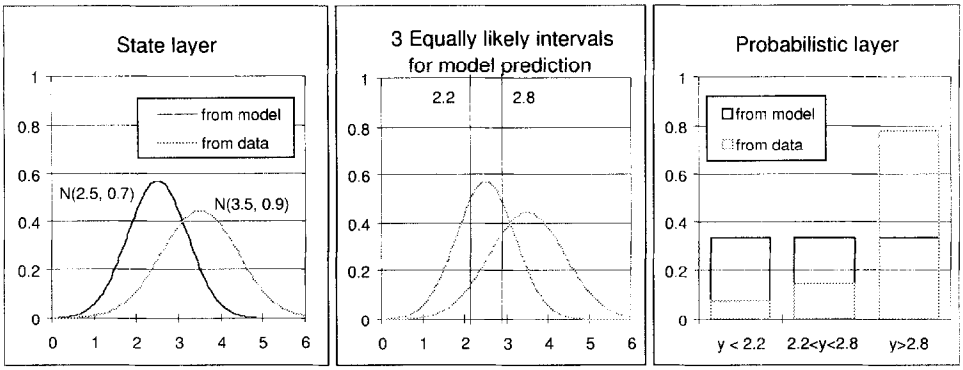

Fig. 2. Discrete approximation of an observation.

- An evidence vector $\boldsymbol{e}_i$, with probabilities $e_1, \ldots, e_k$ assigned to k outcomes. In the present study, the number of outcomes is three. The evidence vector transmits external information (data, targets, etc.) to the model.
- A posterior probability distribution $\mathbf{Bel}_i$.

The prior probabilities assigned to the outcomes are updated with information linked from other parts of the network, yielding the posterior probability distribution.

Links. A probabilistic link (uncertainty link) transfers information from one node to another. It is defined as the link matrix $\boldsymbol{M}_{i|j}$ between two variables i and j, denoting the conditional probability of i given j. In the simplest case of a unidirectional chain, the link matrix equals a Markov chain state transition matrix.

Since the probabilistic layer parallel to the deterministic equations describes their structural uncertainty, the distribution of i should preserve all the moments of the distribution of j (expected value, skewness, kurtosis) except variance. It should be increased correspondingly to the amount of structural uncertainty. It is often practical to give the strength of each link using a single parameter instead of inserting values for each matrix element separately. The following approach fulfills the moment requirements stated above.

The link strength parameter is denoted as $\eta_{j|i}$, $i \neq j$, or just as η. $\eta_{j|i} \in [-1, 1]$. A symmetric, $k \times k$ link matrix $\boldsymbol{M}_{j|i}$ is constructed as a function of $\eta_{j|i}$, which is now used as an input:

$$m_{q,r} = \frac{1}{k} + \eta\left(1 - \frac{1}{k}\right), \quad \eta \geqslant 0\colon q = r = 1, \ldots, k;\ \eta < 0\colon q = k - r,$$
$$r = 1, \ldots, k, \tag{13a}$$

$$m_{q,r} = \frac{1}{k-1}\left[1 - \left[\frac{1}{k} + \eta\left(1 - \frac{1}{k}\right)\right]\right], \quad \eta \geqslant 0\colon q \neq r;\ \eta < 0\colon q \neq k - r. \tag{13b}$$

The sum of the absolute values of the link strength parameters of all the links leaving a node is not allowed to exceed 1. The same applies to all links entering a node. This certifies that no evidence is counted twice in the network. The sum of the absolute values

of link strengths to a node indicates the node's level of dependency of the other nodes in the network. For completely independent variables, it is 0, and for completely dependent ones it is 1.

Network propagation. Two independent likelihood messages, called π and λ, are computed. The updated belief is obtained as the convolution product of these messages and the prior belief. This approach does not update messages in cases where the propagation direction is changed. Computationally, the two propagation directions are symmetric.

When propagating the π messages, all messages coming to a node, say j, from another node, say i, are denoted by $\boldsymbol{p}_{j|i}$ and messages leaving node i are denoted by $\boldsymbol{\pi}_i$. For any node j, preconditioned by any node i $(i < j)$:

$$\boldsymbol{p}_{j|i} = \boldsymbol{M}_{j|i}\boldsymbol{\pi}_i. \tag{14}$$

The likelihood vectors $\boldsymbol{p}_{j|i}$ and $\boldsymbol{\pi}_i$ consist of the following elements:

$$\boldsymbol{\pi}_i = \begin{bmatrix} \pi_i^1 \\ \pi_i^2 \\ \pi_i^3 \end{bmatrix} \quad \text{and} \quad \boldsymbol{p}_i = \begin{bmatrix} p_{j|i}^1 \\ p_{j|i}^2 \\ p_{j|i}^3 \end{bmatrix}. \tag{15}$$

For elements r, the π_i^r message is the scaled vector product (joint distribution) of the message $\pi_{i|1\ldots i-1}^r$ and the evidence e_i^r.

$$\pi_i^r = \pi_{i|1\ldots i}^r = \alpha e_i^r \pi_{i|1\ldots i-1}^r \tag{16}$$

where α is a scaling constant, scaling the sum of the k vector elements of $\boldsymbol{\pi}_i$ to unity. The incoming message $\boldsymbol{\pi}_{i|1\ldots i-1}$ is the joint distribution of all the messages, $\boldsymbol{p}_{i|1}$ to $\boldsymbol{p}_{i|i-1}$, from the node's $i-1$ predecessors:

$$\pi_{i|1\ldots i-1}^r = \prod_{k=1}^{i-1} p_{i|k}^r. \tag{17}$$

Starting from the first node, the $\boldsymbol{p}_{1|0} = \mathbf{1}$ and $\boldsymbol{\pi}_1 = \boldsymbol{e}_1$, $\boldsymbol{p}_{2|0,1} = \boldsymbol{M}_{2|1}\boldsymbol{\pi}_1$ and so on.

The direction is reverse in the λ messages. The rest is computationally similar. All messages coming to node i from node j are denoted by $\boldsymbol{l}_{i|j}$ and messages leaving the node j are denoted by $\boldsymbol{\lambda}_j$. For any node i, preconditioned by any node j, with $i < j$.

$$\boldsymbol{l}_{i|j} = \boldsymbol{M}_{i|j}\boldsymbol{\lambda}_j. \tag{18}$$

The $\boldsymbol{\lambda}_j$ message is the joint distribution of the message $\boldsymbol{\lambda}_{j|j+1\ldots n}$ and the evidence $\boldsymbol{e}_j$.

$$\lambda_j^r = \lambda_{j|j\ldots n}^r = \beta e_j^r \lambda_{j|j+1\ldots n}^r \tag{19}$$

where β is a scaling constant. The incoming message $\boldsymbol{\lambda}_{j|j+1\ldots n}$ is a convolution of all the messages, $\boldsymbol{l}_{j|j+1}$ to $\boldsymbol{l}_{j|n}$, from the node's $n-j$ successors:

$$\lambda_{j|j+1\ldots n}^r = \prod_{k=j+1}^{n} l_{k|j}^r. \tag{20}$$

For each node j, the posterior belief distributions $\mathbf{Bel}_j$ can now be calculated on the basis of the prior distribution $\boldsymbol{e}_j$, updating it with the information from the sub-network before and after the node, i.e., vectors $\boldsymbol{\pi}_{j|1\ldots j-1}$ and $\boldsymbol{\lambda}_{j|j+1\ldots n}$, respectively:

$$Bel_j^r = \gamma \pi_{j|1\ldots j-1}^r e_j^r \lambda_{j|j+1\ldots n} \tag{21a}$$

where γ is a scaling constant. The same equation can be written as a vector product of the two likelihood messages and the evidence vector:

$$\mathbf{Bel}_j = \gamma \boldsymbol{\pi}_{j|1\ldots j-1} \bullet \boldsymbol{e}_j \bullet \boldsymbol{\lambda}_{j|j+1\ldots n}. \tag{21b}$$

3.4. Information from probabilistic layer to state layer

In the approach proposed, there are two different paths of information from the probabilistic layer to the state layer:

- The link strength parameter η is involved in the propagation of uncertainty (9).
- The deviations between prior and posterior distributions give important diagnostic information about the model. In parameter estimation or other adjustment of the model to fulfill given targets, the posteriors are iterated to make them uniform distributions.

The suggested quadratic/linear iteration scheme (22) providing rapid convergence is based on comparison of the probabilities of the different outcomes of a control variable. They are iterated to be equal to one another.

$$\mu_i^* = \mu_i + a \cdot \mathrm{cv}_i \cdot \left(Bel_y^1 - Bel_y^3\right) \left|Bel_y^1 - Bel_y^3\right|, \tag{22a}$$

$$\eta_i^* = \eta_i + b \cdot \left(Bel_y^2 - 1/k\right), \tag{22b}$$

where a and b are convergence parameters, Bel_y^r is the posterior probability of outcome r, k is the number of outcomes, μ_i is the mean of the prior distribution of node i (a control variable), η_i is the estimated link strength, and * refers to an updated iteration value. This iteration scheme was found experimentally to be markedly more rapid and practical than parametric approaches such as t-test based iteration.

4. Numerical examples

4.1. Two-directional propagation of uncertainty

The two-directional uncertainty scheme of belief networks is illustrated with the following example which, for simplicity, has no state layer. The example comes from fish stock assessment that are needed to impose proper fisheries restrictions. Extensive data collection from nature is most often out of the question due to high costs, and indirect data are typically used. This type of data tends to be corrupted by many types of biases. Decisions on allowable catches are needed regularly, typically on an annual basis.

The simplest possible model for the system includes two mutually dependent variables: fish stock and fish catch per fishing unit (e.g., one fishing night; Fig. 3). This dependency is usually used in assessment of both variables. There are several ways of obtaining

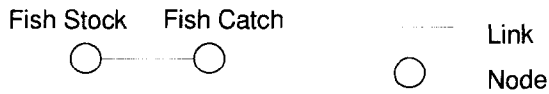

Fig. 3. Structure of the example model. In general, the links are two-directional.

independent information on them. Here, fish stock assessment is based on catch estimates and the number of returned taggings, and the catch assessment on stock estimates and taxation records of professional fishermen or enterprises. The outcomes of both variables are, in relation to the previous year, say a 30% decrease, an unchanged level, and a 30% increase.

A methodologically interesting question arises from the fact that, in the scale under consideration, fish stock can be understood as the cause and fish catch as the effect. Assessment from cause to effect and vice versa is clearly a strength in any environmental and resource management task. In a longer time frame, over several years, there is also a feedback from fish catch to fish stock.

The following notation is used: $\boldsymbol{e}_{\text{stock}}$ is the information from returned taggings, $\boldsymbol{e}_{\text{catch}}$ is the one from taxation records, π is the likelihood message from fish stock to fish catch, λ is the one from fish catch to fish stock, $\boldsymbol{M}$ is the link matrix which is equal in both directions, and α and β are scaling parameters. Now, we obtain the posteriors of the elements r of variables $\mathbf{Bel}_{\text{stock}}$ and $\mathbf{Bel}_{\text{catch}}$ by

$$Bel^r_{\text{stock}} = P\left(\text{stock}^r \mid e^r_{\text{catch}}\right) = \alpha P\left(\text{stock}^r\right)\lambda^r = \alpha e^r_{\text{stock}}\lambda^r,$$
$$Bel^r_{\text{catch}} = P\left(\text{catch}^r \mid e^r_{\text{stock}}\right) = \beta P\left(\text{catch}^r\right)\pi^r = \beta e^r_{\text{catch}}\pi^r.$$

The messages $\boldsymbol{\pi}$ and $\boldsymbol{\lambda}$ are

$$\boldsymbol{\pi} = \boldsymbol{M}\boldsymbol{e}_{\text{stock}}, \qquad \boldsymbol{\lambda} = \boldsymbol{M}\boldsymbol{e}_{\text{catch}}.$$

Examine now the propagation scheme with four numerical cases.

(a) The link matrix is as given in Fig. 4, and information from returned taggings is $\boldsymbol{e}_{\text{stock}} = (0.1, 0.3, 0.6)^{\text{T}}$, implying that the stock is likely to grow.
(b) Assume that information, instead of stock, exists on catch only. Now $\boldsymbol{e}_{\text{catch}} = (0.8, 0.15, 0.05)^{\text{T}}$.
(c) All the above information is simultaneously available. This controversial information forces both the belief vectors close to noninformative ones.
(d) The evidence vectors support one another.

This results a higher belief on increasing stocks and catches than the evidence vectors alone would suggest.

4.2. Two-layered model including deterministic dependencies

Let us elaborate the above example further to demonstrate the use of deterministic equations between two variables. Such highly aggregated equations are often used in practice due to convention, transparency requirements, resource constraints etc., although they are known not to describe the phenomenon under concern with full certainty. The

FISH STOCK

(a)

y	e	Bel	λ
- 30%	0.1	**0.1**	0.33
Same	0.3	**0.3**	0.33
+ 30%	0.6	**0.6**	0.33

M		
0.73	0.13	0.13
0.13	0.73	0.13
0.13	0.13	0.73

FISH CATCH

π	Bel	e	y
0.19	**0.19**	0.33	- 30%
0.31	**0.31**	0.33	Same
0.49	**0.49**	0.33	+ 30%

(b)

y	e	Bel	λ
- 30%	0.33	**0.61**	0.61
Same	0.33	**0.22**	0.22
+ 30%	0.33	**0.16**	0.16

M		
0.73	0.13	0.13
0.13	0.73	0.13
0.13	0.13	0.73

π	Bel	e	y
0.33	**0.8**	0.8	- 30%
0.33	**0.15**	0.15	Same
0.33	**0.05**	0.05	+ 30%

(c)

y	e	Bel	λ
- 30%	0.1	**0.27**	0.61
Same	0.3	**0.3**	0.22
+ 30%	0.6	**0.43**	0.16

M		
0.73	0.13	0.13
0.13	0.73	0.13
0.13	0.13	0.73

π	Bel	e	y
0.19	**0.68**	0.8	- 30%
0.31	**0.21**	0.15	Same
0.49	**0.11**	0.05	+ 30%

(d)

y	e	Bel	λ
- 30%	0.1	**0.07**	0.25
Same	0.3	**0.25**	0.31
+ 30%	0.6	**0.69**	0.43

M		
0.73	0.13	0.13
0.13	0.73	0.13
0.13	0.13	0.73

π	Bel	e	y
0.19	**0.1**	0.2	- 30%
0.31	**0.25**	0.3	Same
0.49	**0.65**	0.5	+ 30%

Fig. 4. Propagation of fish stock information to fish catch (a) and vice versa (b); impacts of controversial (c) and mutually supporting (d) information on posteriors.

structural (causal) uncertainty involved in the model can be modeled using a belief network in the following manner.

Instead of using outcomes y_i such as (−30%, Same, +30%), we use now numerical quantities, e.g., for $y_{\text{catch}} = (70\,000, 100\,000, 130\,000)^{\mathrm{T}}$. The stock assessment given the catch can be done in many ways. One of the standards is the following stock equation:

$$y_{\text{stock}} = y_{\text{catch}} \frac{m + F}{F(1 - \mathrm{e}^{-m-F})}$$

where F is fishing mortality rate and m is natural mortality rate. Use, for example, the rates: $F = 0.3$ and $m = 0.1$. Using the stock equation, $y_{\text{stock}} = (283\,000, 404\,000, 526\,000)^{\mathrm{T}}$. Taking now the evidence vectors and the link matrix from Fig. 4(a), the resulting model is as shown in Fig. 5(a). For further illustration, Fig. 5(b) shows the case in which the rate parameter values have been changed to $F = 0.5$ and $m = 0.2$, in Fig. 5(c) $y_{\text{catch}} = (10\,000, 20\,000, 30\,000)^{\mathrm{T}}$, and in Fig. 5(d) the link matrix has been changed to imply a weaker dependency between the variables, and $\boldsymbol{e}_{\text{catch}} = (0.8, 0.15, 0.05)^{\mathrm{T}}$.

4.3. Parameter estimation by uncertainty balance

This example contains, besides the state and the probabilistic layers, also targets (observations), and decision variables (parameters). To define the state layer, consider the following linear model

$$y_{i+1} = a y_i, \quad i = 1, 2, 3,$$

	FISH STOCK							FISH CATCH			
(a)	y	e	Bel	λ		M		π	Bel	e	y
	283 103	0.1	**0.1**	0.33	0.73	0.13	0.13	0.19	**0.19**	0.33	*70 000*
	404 433	0.3	**0.3**	0.33	0.13	0.73	0.13	0.31	**0.31**	0.33	*100 000*
	525 762	0.6	**0.6**	0.33	0.13	0.13	0.73	0.49	**0.49**	0.33	*130 000*
(b)	y	e	Bel	λ		M		π	Bel	e	y
	194 671	0.1	**0.1**	0.33	0.73	0.13	0.13	0.19	**0.19**	0.33	*70 000*
	278 101	0.3	**0.3**	0.33	0.13	0.73	0.13	0.31	**0.31**	0.33	*100 000*
	361 531	0.6	**0.6**	0.33	0.13	0.13	0.73	0.49	**0.49**	0.33	*130 000*
(c)	y	e	Bel	λ		M		π	Bel	e	y
	27 810	0.1	**0.1**	0.33	0.73	0.13	0.13	0.19	**0.19**	0.33	*10 000*
	55 620	0.3	**0.3**	0.33	0.13	0.73	0.13	0.31	**0.31**	0.33	*20 000*
	83 430	0.6	**0.6**	0.33	0.13	0.13	0.73	0.49	**0.49**	0.33	*30 000*
(d)	y	e	Bel	λ		M		π	Bel	e	y
	27 810	0.1	**0.14**	0.43	0.47	0.27	0.27	0.29	**0.77**	0.8	*10 000*
	55 620	0.3	**0.3**	0.3	0.27	0.47	0.27	0.33	**0.16**	0.15	*20 000*
	83 430	0.6	**0.56**	0.28	0.27	0.27	0.47	0.39	**0.06**	0.05	*30 000*

Fig. 5. Two-layered propagation examples. Figures in italics: the outcome layer, other figures: the probabilistic layer.

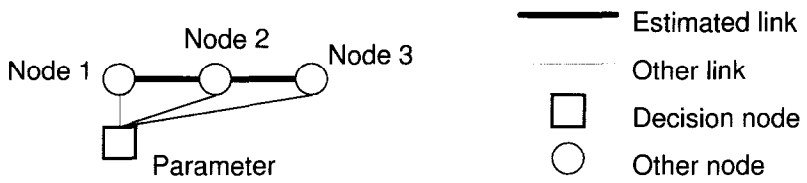

Fig. 6. Structure of the example model.

where y_i is the model prediction of an observed variable e_i^* at point i, and a is a parameter. All these variables are normally distributed. The tasks are to estimate the expected value of the parameter a, and to estimate the model's structural uncertainty (= link strengths). These estimates are based on the three observations e_1^*, e_2^*, e_3^*. Fig. 6 presents the structure of the model, and the Microsoft Excel code is given in Fig. 7.

No external information. In the following, the estimation procedure is illustrated with a numerical example, and the propagation scheme is calculated stepwise. In the first step, a model is present with no observations. As it now includes the state layer and the probabilistic layer, it takes the form shown in Fig. 8. Due to the discrete approximation principle (Fig. 2), all distributions are uniform if no external information is there. Therefore, changes introduced in parameters or initial states introduce no changes in the

	A	B	C	D	E
1			Node 1	Node 2	Node 3
2	Model output	y_{-}	3	=B41*C2	=B41*D2
3		cv_{model}	=C6/SQRT(B14)	=C3	=D3
5	Observation	e^{*}_{1}	2.5		
6		cv_{data}	0.3	=C6	=C6
8		e_1	π_1	Bel_1	$l_{1|2,3}$
9	Node 1	=IF(C5='',0.333,normstodisc(C2,C3,C5,C6))	=B9/SUM(B9:B11)	=C9*E9/(C9*E9+C10*E10+C11*E11)	=MMULT(C13:E15,C23:C25)
10		=IF(C5='',0.333,normstodisc(C2,C3,C5,C6))	=B10/SUM(B9:B11)	=C10*E10/(C9*E9+C10*E10+C11*E11)	=MMULT(C13:E15,C23:C25)
11		=IF(C5='',0.333,normstodisc(C2,C3,C5,C6))	=B11/SUM(B9:B11)	=C11*E11/(C9*E9+C10*E10+C11*E11)	=MMULT(C13:E15,C23:C25)
13	Link matrix	$\eta_{1-2} = \eta_{2|1}$	=0.333+0.666*B14	=C14	=C14
14	$M_{1|2} = M_{2|1}$	=iterated_link_strength	=(1-C13)/2	=C13	=C14
15			=C14	=C14	=C13
17		e_2	$\pi_{2|1}$	p_2	
18	Node 2	=IF(D5='',0.333,normstodisc(D2,D3,D5,D6))	=MMULT(C13:E15,C9:C11)	=C18*B18	
19		=IF(D5='',0.333,normstodisc(D2,D3,D5,D6))	=MMULT(C13:E15,C9:C11)	=C19*B19	
20		=IF(D5='',0.333,normstodisc(D2,D3,D5,D6))	=MMULT(C13:E15,C9:C11)	=C20*B20	
22		Bel_2	l_2	$\lambda_{2|3}$	
23		=D18*D23/(D18*D23+D19*D24+D20*D25)	=D23*B18	=MMULT(C27:E29,E32:E34)	
24		=D19*D24/(D18*D23+D19*D24+D20*D25)	=D24*B19	=MMULT(C27:E29,E32:E34)	
25		=D20*D25/(D18*D23+D19*D24+D20*D25)	=D25*B20	=MMULT(C27:E29,E32:E34)	
27	Link matrix	$\eta_{2-3} = \eta_{3|2}$	=0.333+0.666*B28	=C28	=C28
28	$M_{2|3} = M_{3|2}$	=B14	=(1-C27)/2	=C27	=C28
29			=C28	=C28	=C27
31		e_3	$p_{3|1,2}$	Bel_3	λ_3
32	Node 3	=IF(E5='',0.333,normstodisc(E2,E3,E5,E6))	=MMULT(C27:E29,D18:D20)	=C32*E32/(C32*E32+C33*E33+C34*E34)	=B32/SUM(B32:B34)
33		=IF(E5='',0.333,normstodisc(E2,E3,E5,E6))	=MMULT(C27:E29,D18:D20)	=C33*E33/(C32*E32+C33*E33+C34*E34)	=B33/SUM(B32:B34)
34		=IF(E5='',0.333,normstodisc(E2,E3,E5,E6))	=MMULT(C27:E29,D18:D20)	=C34*E34/(C32*E32+C33*E33+C34*E34)	=B34/SUM(B32:B34)
36	Link matrix	$\eta_{4|1,2,3}$	=0.333+0.666*B37	=C37	=C37
37	$M_{4|123}$	1	=(1-C36)/2	=C36	=C37
38			=C37	=C37	=C36
40		a	cv_{model}		
41	Parameter	=iterated_parameter_a	=C3		
43		e_4	π_4	Bel_4	
44	Node 4	0.333	=B44/SUM(B44:B46)	=C44*B49/(C44*B49+C45*B50+C46*B51)	
45		0.333	=B45/SUM(B44:B46)	=C45*B50/(C44*B49+C45*B50+C46*B51)	
46		0.333	=B46/SUM(B44:B46)	=C46*B51/(C44*B49+C45*B50+C46*B51)	
48		pl_4	pl_{41}	pl_{42}	pl_{43}
49		=C49*D49*E49/(C49*D49*E49+C50*D50*E50+C51*D51*E51)	=MMULT(C36:E38,D9:D11)	=MMULT(C36:E38,B23:B25)	=MMULT(C36:E38,D32:D34)
50		=C50*D50*E50/(C49*D49*E49+C50*D50*E50+C51*D51*E51)	=MMULT(C36:E38,D9:D11)	=MMULT(C36:E38,B23:B25)	=MMULT(C36:E38,D32:D34)
51		=C51*D51*E51/(C49*D49*E49+C50*D50*E50+C51*D51*E51)	=MMULT(C36:E38,D9:D11)	=MMULT(C36:E38,B23:B25)	=MMULT(C36:E38,D32:D34)

Fig. 7. (Sideways.) Excel code of the parameter estimation example (cf. Fig. 9). The normstodisc function discretizes normal distributions using Eqs. (10)–(12). The iterated values in cells B14 and B41 are obtained using Eq. (22).

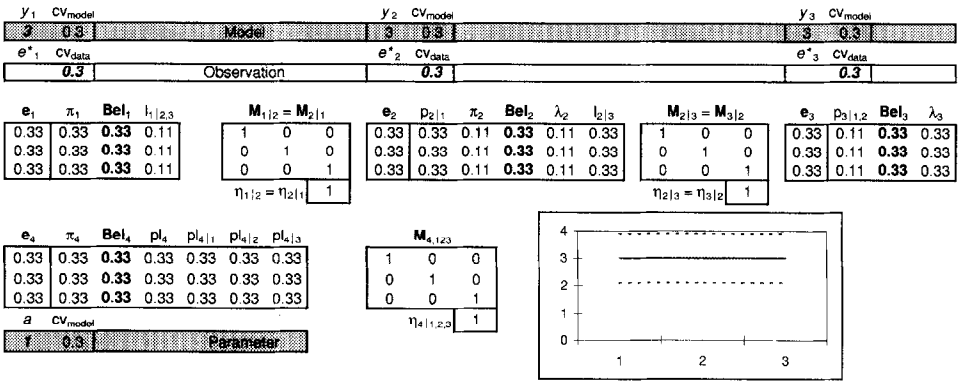

Fig. 8. The model with no external information.

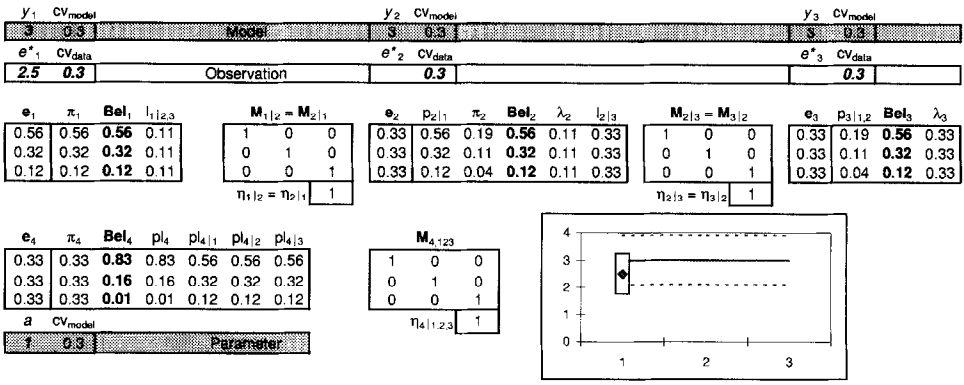

Fig. 9. Propagation of the observation e_1^*. A discrete approximation (evidence vector $\boldsymbol{e}_1$) is made, and it is propagated through the π system.

probabilistic layer. Changes in link strength values change only values in the link matrices, but do not influence any of the probability distributions.

One observation is included. When adding an observation at any node—say, node 1 as an example—the continuous distribution of the observation e_1^* is approximated with a discrete distribution having the same outcomes as were used in the discrete approximation of the model output distribution at node 1 (Eqs. (10)–(12)). The information in $\boldsymbol{e}_1$ is included in the π message, and is now propagated through the network (Fig. 9). Note that the posterior distributions (**Bel**s) now equal the π messages, because there is no information coming up to the λ system. The nonuniform distributions imply that there is also other information available besides the model, including $\mathbf{Bel}_4$. This feature will be used later in parameter estimation.

More than one observation. Now, add an observation into node 3 (Fig. 10). A discrete approximation is made to the distribution of e_3^*, and the information is propagated through

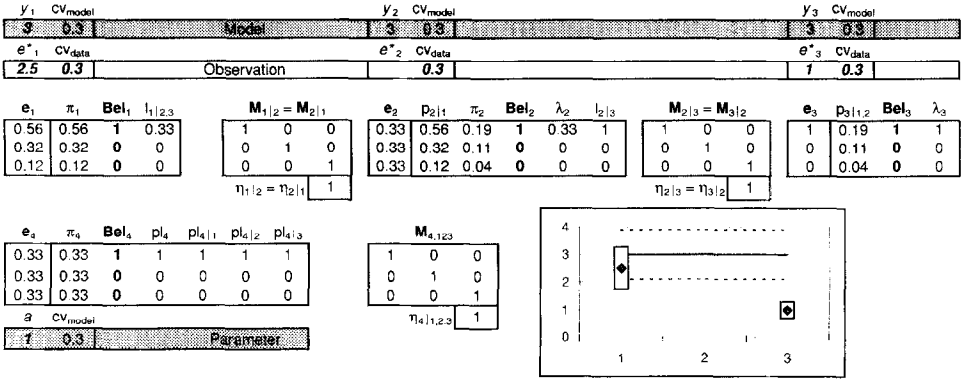

Fig. 10. Propagation of the observation e_3^*. A discrete approximation (evidence vector e_3) is made, and it is propagated through the λ system.

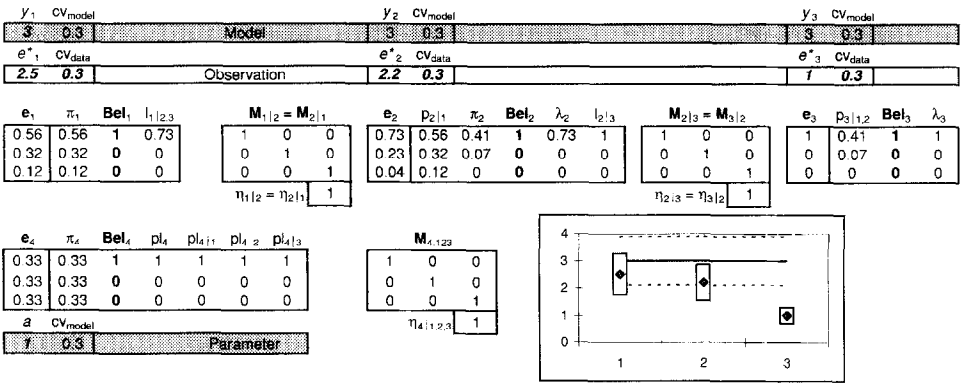

Fig. 11. Propagation of the observation e_2^*. A discrete approximation (evidence vector e_2) is made, and it is propagated through the π system to the direction of node 1, and through the λ system to the direction of node 3.

the network. Correspondingly, we can add an observation to node 2 (Fig. 11). Note that the **Bel**s are no longer equal to either the π or the λ messages, but their scaled vector product. The posterior of parameter $\mathbf{Bel}_4$ is again updated.

Parameter estimation. This step estimates (1) a value to the parameter and (2) the link strengths between nodes 1 and 2, and nodes 2 and 3 (Fig. 12). The principle used can also be applied to many other optimization tasks, as will be shown in the river example later on. The idea is to obtain such values to the parameter and the link strength that $\mathbf{Bel}_4$ becomes uniform.

Fig. 13 gives a set of examples of possible distributions of $\mathbf{Bel}_4$, and of the inference that can be made on the basis of such distributions. Note that when either a parameter value, link strength value, or observed value is changed, the probability values in the evidence vectors are also changed, because the outcome distributions change.

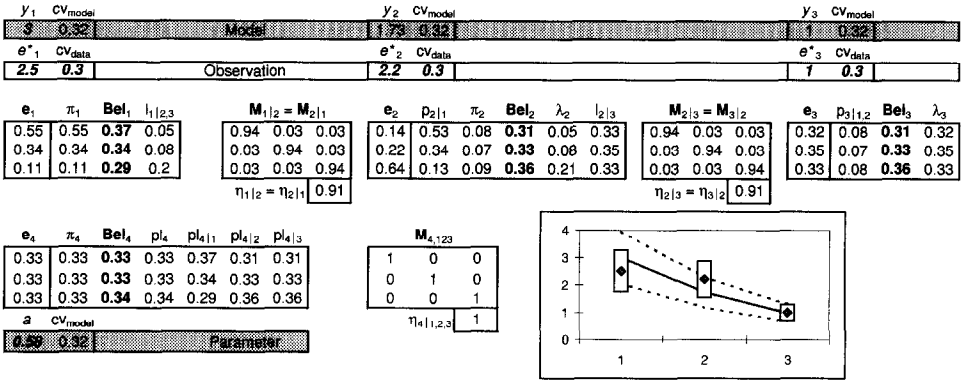

Fig. 12. The model after iteration of $\mathbf{Bel}_4$ to be a uniform distribution.

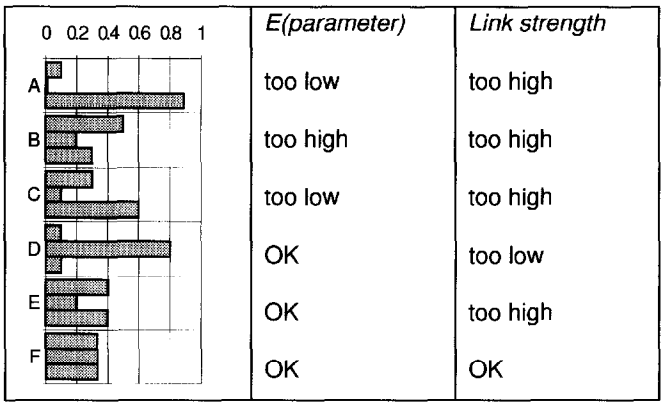

Fig. 13. Some example posterior distributions of the parameter ($\mathbf{Bel}_4$) and the inference based on these types of distributions.

5. A river quality management case

5.1. The management problem and the watershed

This example deals with cost-effective upgrading of wastewater treatment plants in a watershed on the basis of ambient water quality criteria. It represents a classical river basin management problem [32]. Priorization problems of this character are of particular interest to funding organizations and government agencies in countries short of capital. The example was generated within the context of a comprehensive priorization study including several former socialist countries in Europe [29]. They are in the midst of a very rapid and profound transition process, affecting all sectors of the societies, including water quality management. Previously, the integration of ambient and effluent monitoring has been low. At present, the industry is undergoing considerable change, and past water quality data are of limited validity, yet there is a pressing need for improving water pollution

control. The scarcity of capital suggests the policy of gradual upgrading of wastewater treatment on a cost-effective basis [28]. A more detailed documentation of the study is given elsewhere [34].

A hypothetical watershed is used with ten municipal wastewater treatment plants to be upgraded to improve the river water quality. Each plant discharges the effluent into a different tributary. The impact of different ambient water quality criteria and diverse investment levels are studied under the precepts of cost-effective prioritization of upgrading levels at the plants. A variety of treatment alternatives is available, ranging from no treatment (level 0), chemical treatment (1) to more advanced solutions [29,34]. Initially, all the plants here are at level 1.

5.2. A probabilistic river model

Based on the results of the comprehensive water quality management study of the Nitra River Basin, Slovakia [14,29], an extended Streeter–Phelps model with three state variables was chosen for this study. The state variables are dissolved oxygen (DO), biological oxygen demand (BOD) and ammonium (NH4). The state equations that describe the steady-state evolution of the river water quality are:

$$\frac{\mathrm{dBOD}}{\mathrm{d}\tau} = -k_1\mathrm{BOD},$$

$$\frac{\mathrm{dNH4}}{\mathrm{d}\tau} = -k_2\mathrm{NH4},$$

$$\frac{\mathrm{dDO}}{\mathrm{d}\tau} = -k_1\mathrm{BOD} - k_2\mathrm{NH4} + k_3(\mathrm{DO_s} - \mathrm{DO})$$

where τ is water travel time, $\mathrm{DO_s}$ is the saturation concentration of DO in water, and k_i are three rate parameters that are estimated: k_1 is BOD oxygenation, k_2 is NH4 oxygenation, and k_3 is reaeration rate. The unit of state equations is mass per time. These equations are analytically solvable, and their analytical solutions are used as the state layer. The probabilistic layer is based on a network corresponding the river topology (Fig. 14(a)). State variables and parameters are represented as belief network nodes. Evidential information for states is obtained from field measurements.

The analysis is divided into two subsequent phases, at both of which the uncertainty balance iteration approach is used. The same model including the two layers is used, but the targets, decision variables, and estimated link strengths are different (Table 1, Figs. 14(b) and 14(c)). First, the parameter estimation is performed, in which the mean values at the state layer are iterated to equal the posteriors. The link strengths of the links

Table 1
Definition of decision (control) variables and targets in the diagnostic and in the policy analysis parts of the study

	Diagnosis	*Policy model*
Decision variables	Parameters	Dischargers
Targets	Observed water quality	Water quality targets. Target costs

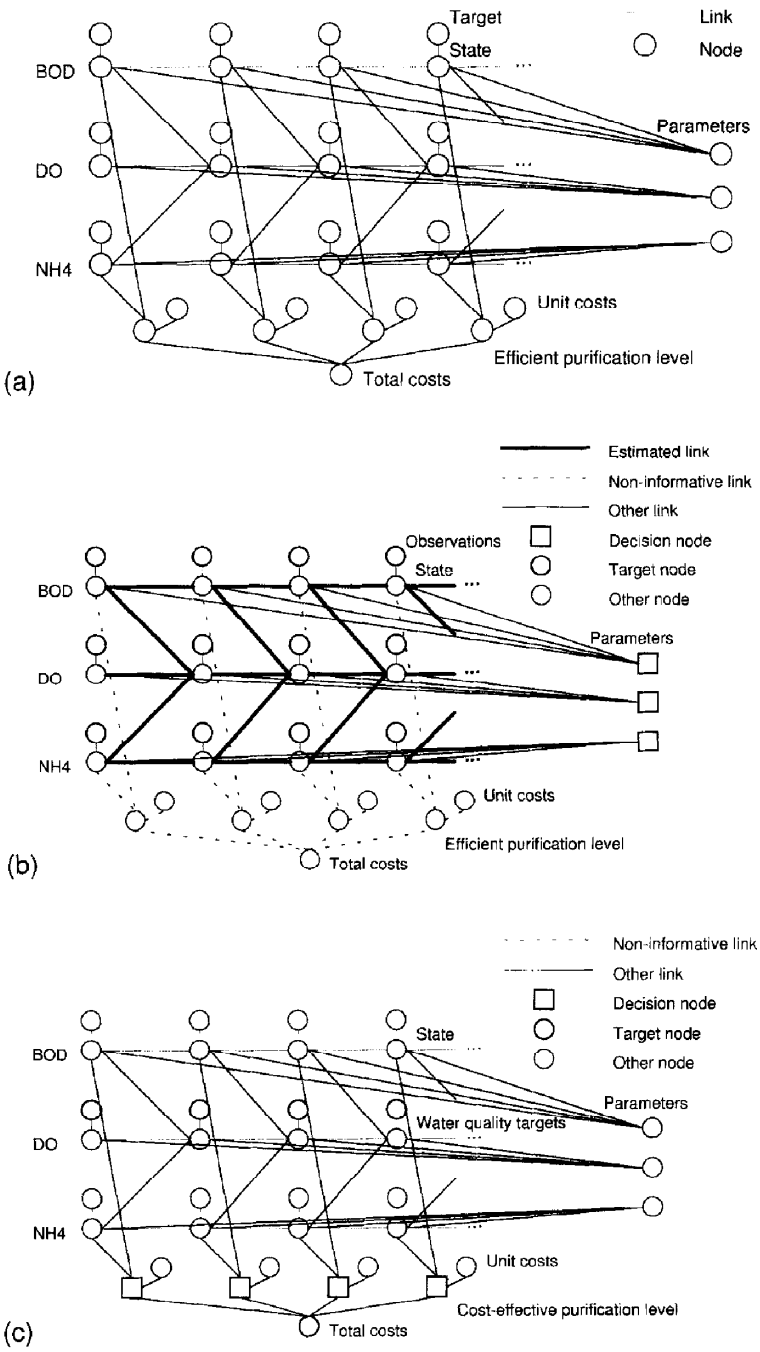

Fig. 14. Probabilistic layer of the river model. (a) configuration, (b) the diagnostic phase: observations are targets, and model parameters are control variables are, and (c) the policy analysis phase: water quality criteria and total costs are targets, and purification levels are control variables.

shown in Fig. 14(b) are estimated; their values show the structural uncertainties of the state equations.

The second phase consists of finding the most cost-effective solutions for river water quality management, taking into account the water quality targets for the river and the costs involved. Now, different treatment levels are used as decision variables (instead of parameters at the previous phase), link strengths are not estimated, and water quality targets together with the target cost level are used as targets (versus observations at the previous phase).

The definition of variables can be changed in the course of the analysis due to the two-directional uncertainty propagation scheme. At the diagnostic phase, both downstream and upstream (π and λ, respectively) messages are used. In the policy analysis phase, only the λ message going upstream is used. This is because in the diagnosis, all the data and model predictions are iterated to meet the balance, hence both propagation directions are used. In the policy analysis phase, the targets influence only the treatment plants downstream of the point at which a target is set. When detecting a deviation between target and model prediction, the message induced is propagated upstream all the way to the posterior distributions of the treatment plant purification levels. This provides a basis for iteration similar to that in parameter estimation.

5.3. Model versus data: illustration of the approach

The model calibration for the hypothetical data is shown in Fig. 15. Take an example of the propagation of evidential information (observations) in one of the model equations, say

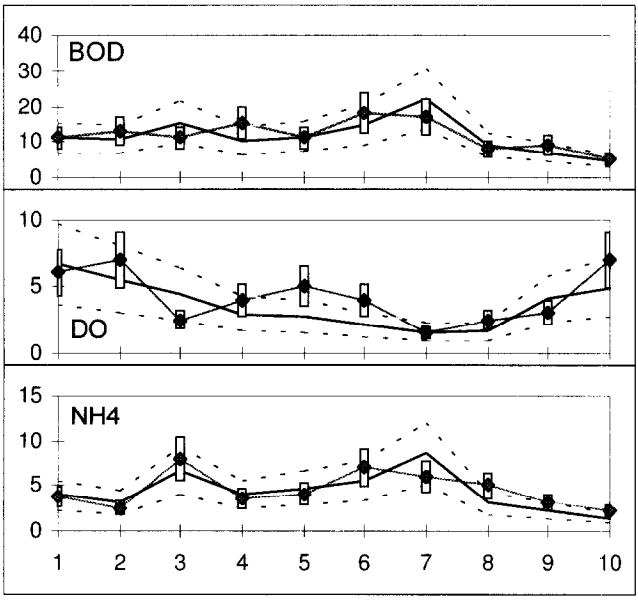

Fig. 15. The nominal case. Calibrated link strengths are 0.83 for BOD, 0.87 for DO, and 0.81 for NH4. The dots stand for observations, and the solid line for the calculated level. 90% confidence intervals are also shown.

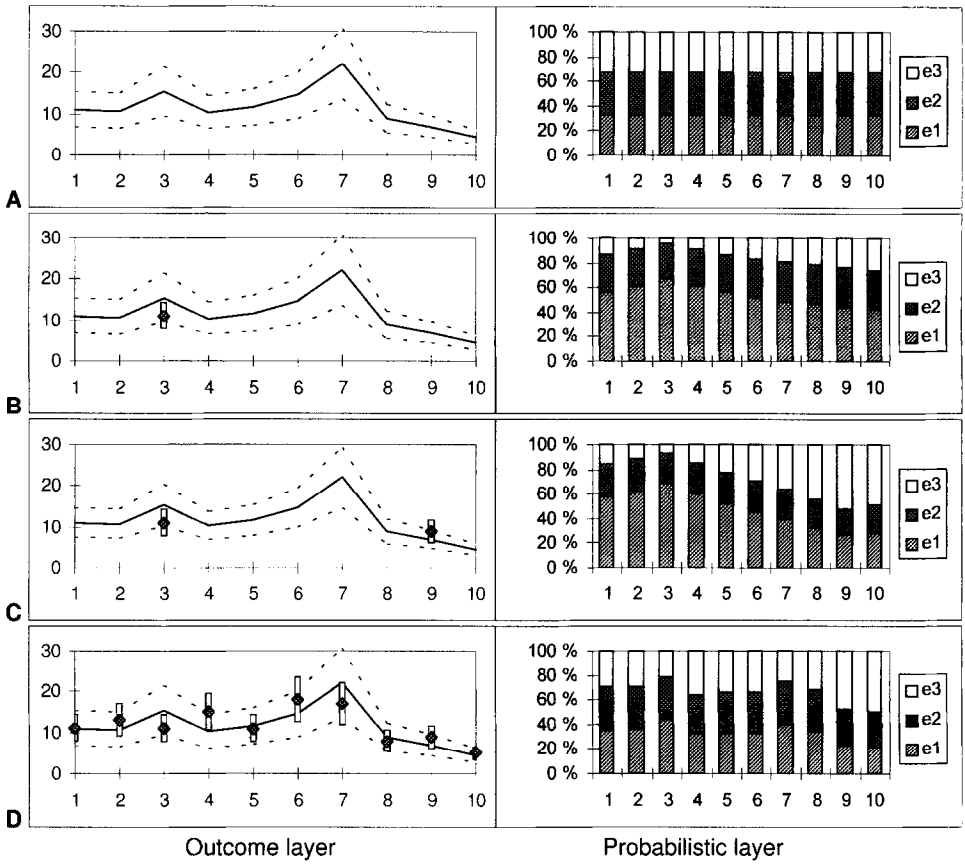

Fig. 16. Propagation of observed information in the probabilistic layer. The columns on the right show the posterior distributions (**Bel**s) of the model prediction at different points.

BOD (Fig. 16). We want to know how well the model is modeling our system and we want the model/system correspondence to be as good as possible. The system reference consists in this case of observations. In case A, there are no observations, and the probabilistic layer consists of uniform distributions, and there is no information on the success in our modeling task. In case B, there is one observation available, and its probability distribution is discretized according to the proposed procedure (Eqs. (10)–(12); Fig. 2). This vector is used as an evidence vector, and this information is propagated throughout the net (Fig. 16). Case C includes one more observation, which is again discretized and fed into the net. Case D includes a series of observations. In cases B, C, and D, the parameter estimation is based on the iteration of the joint distribution (scaled vector product) of the **Bel**s to a uniform distribution. The more differences there are in the column heights on the right-hand figures, the more misfit there is between model and data. The better the model fit, the lower is its structural uncertainty.

We next demonstrate the influence of prior uncertainties and model fit on posterior uncertainties (Table 2, Fig. 17). The data may have a low uncertainty, but the model fit is poor. In such a case (case A), the link strengths become low, structural uncertainty of the model becomes high, and the prediction highly uncertain. If the data have high uncertainty and the fit is poor, the link strengths may still be high, but prediction remains uncertain (case B). If the data have low uncertainty and the mode fit is good, then the link strengths are high and prediction has low uncertainty (case C). If the data have high uncertainty and the model fit is good, then the link strengths are high and prediction has high uncertainty (case D).

The issue becomes more complex if the model has more than one state equation, as is the case with the river model. Here, the state equations are inter-linked so that the parameters for the BOD and NH4 equations must be estimated first, and thereafter the DO equation is in turn. Success in the BOD prediction strongly depends on success in predicting the BOD and NH4 concentrations along the river. However, it may often happen that the

Table 2
Different typical combinations of prior information and their influence on posterior information (cf. Fig. 17)

Case	*Priors*		*Posteriors*	
	E(data) versus E(model prediction)	*Uncertainty(data)*	*Link strengths*	*Uncertainty (Model prediction)*
A	low accordance	low	low	high
B	low accordance	high	high	high
C	high accordance	low	high	low
D	high accordance	high	high	high

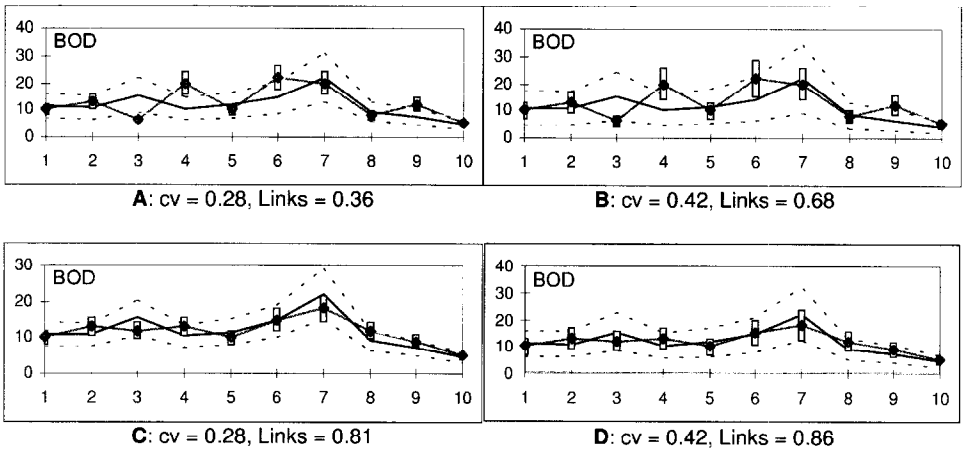

Fig. 17. Example with the BOD equation: The influence of uncertainties and controversies in prior information on uncertainties in the posterior information (cf. Table 2).

uncertainties within the BOD and NH4 data are much higher than those within the DO data, due to analytical accuracy and variability of different substances in nature. Under such conditions, the uncertainties of the BOD and NH4 predictions become high and are propagated throughout the state equations (state layer) to the DO prediction, which also becomes highly uncertain (Fig. 18). This occurs despite good empirical evidence. In such a case, the model structure might no longer be as efficient and another model configuration could be considered, for instance, an empirical model based more strongly on high-quality empirical evidence.

Often, for different reasons, a prefixed set of parameter values is used. These values might be used in standard fashion or when the use of literature values would be considered more adequate than that of empirical parameters. In such a case, the link strength becomes lower than it would be if empirical parameter values were used, unless they were equal. Accordingly, we pay a price for using standard parameter values under conditions of higher uncertainty in prediction.

At the policy analysis phase, the analysis follows the same outline. The largest difference in the example is, however, that link strengths are not estimated at this phase. An interesting phenomenon occurs if the target economic level is set too high compared with the ambient water quality targets. The approach does not find a single solution, because there is looseness in the targets. Either the economic targets should be set lower or the ambient targets should be higher, or both should be done to find a single solution.

To meet the objectives set in the beginning of the example—to provide support in the prioritization of the upgrading of wastewater treatment plants—the model can be used

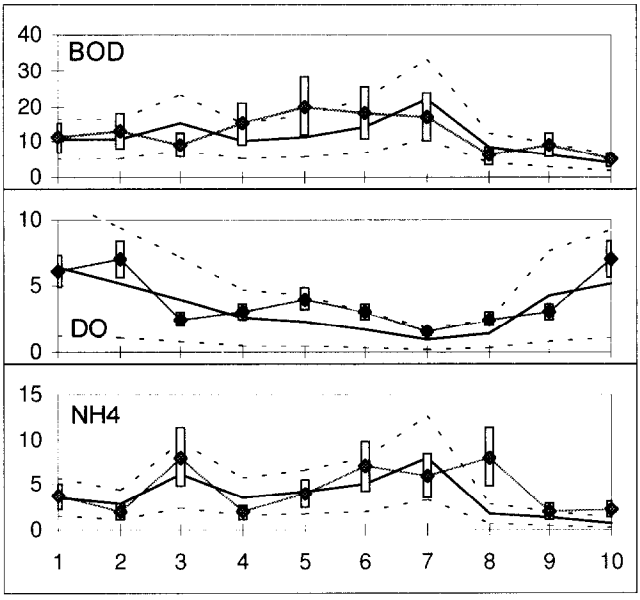

Fig. 18. The DO prediction remains highly uncertain even though DO observations are very accurate, but if BOD and/or NH4 predictions have high uncertainty.

to produce policy scenarios with diverse targets for ambient water quality and varying economic targets. There are many ways of setting water quality targets. The study [34] presents five scenarios:

(1) equal purification level (equity to all dischargers, the prevalent paradigm in Western countries);
(2) equal improvement in all parts of the river;
(3) minimum concentration level in any part of the river;
(4) target(s) at specific point(s) of the river (city, outflow, water intake, recreational area, protected site, etc.); and
(5) target probability levels (frequencies of occurrence; risk averse approach).

Fig. 19 shows an example of these results, comparing two cases from these scenarios (1) and (2). The comparison of conventional, normative strategy, in which all dischargers are imposed to equal purification standards would be more expensive and yield less improvement in river quality than an environment-based approach.

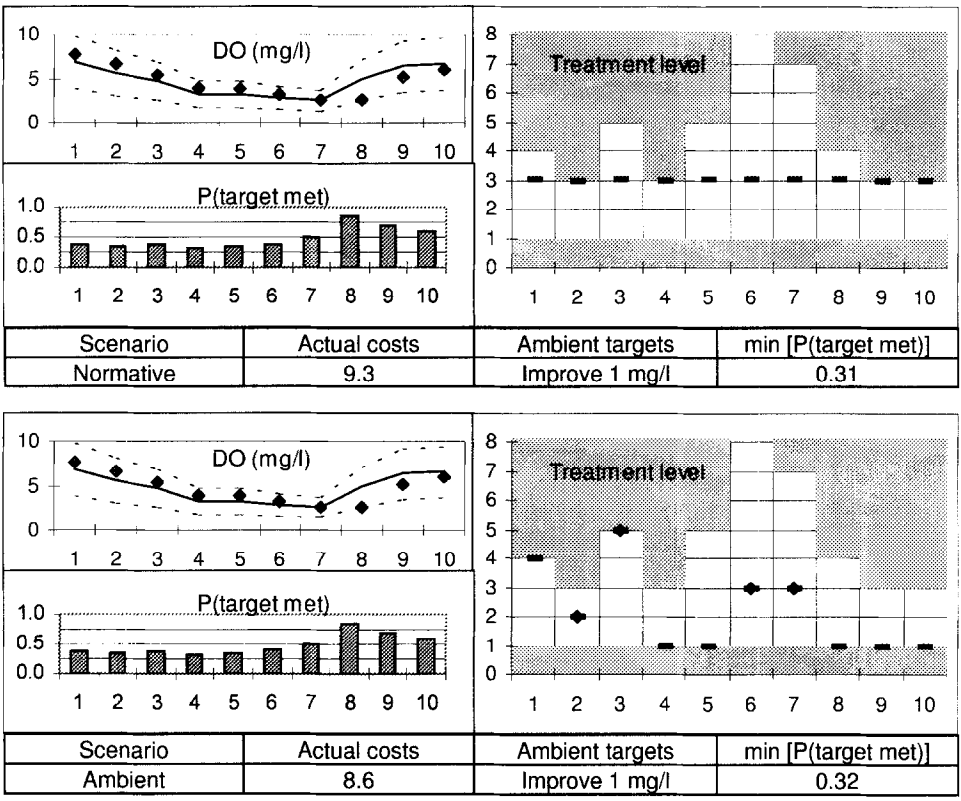

Fig. 19. Sample policy scenarios: comparison of normative (above) and ambient-based (below) results. The latter is more cost-effective than the former. In the upper-left plots, the dots show the target levels, the solid line the optimized level, and the dotted lines indicate its 90% confidence interval.

The most important feature in the results of these scenarios is that normative, equal treatment level at all plants is the most costly way of improving ambient water quality among the studied options, which clearly justifies the problem setting. Another issue is whether the prioritization is workable institutionally and juridically in the countries of the region.

6. Discussion and conclusions

An approach to using belief networks for probabilistic modeling in optimization and parameter estimation is presented. The approach enables updating of uncertainties in different model components interactively. It is therefore remarkably rapid, particularly when compared with conventional approaches requiring off-line simulation runs. Approaches such as Monte Carlo simulation, although practicable in many cases, are time-consuming, noninteractive, and have been criticized as being inaccurate [15]. The major disadvantage of the proposed approach is the relatively labor-intensive computer implementation when compared with conventional simulation approaches, at least at the pilot-study phase documented here.

The proposed approach can be used to detect inconsistencies among different pieces of information in different model components. Possible inconsistency appears as a difference between Bayesian prior and posterior distributions, in a given model component. This feature was used to develop an optimization approach in which prior and posterior distributions of objective functions are iterated to become equal. This can be done by changing the values of control variables and adjusting linking properties in the belief network.

The uncertainty balance approach can handle more than one objective function simultaneously. For instance, in the river basin example, the management optimization included two objectives: target costs and target ambient water quality. The approach finds a compromise (trade-off) between these targets, and can thus be used as a multiobjective optimization approach.

Within environmental and resource management sectors, common practical management models are relatively simple constructs that often can be analytically solved. The use of relatively simple and well-known or easily comprehensible, conceptual models is a great advantage in practical assessment and policy modeling. Transparency and quality assurance are often critical points when striving for the proper, critical attitude and utilization of modeling results. The proposed approach allows consideration of such models as uncertain constructs. The structural uncertainty can be estimated empirically, and the models can be linked and fused with other pieces of probabilistic information.

In the management of natural resources and the environment, the uncertainties are often very high or extreme. In the case of probabilistic models, this means that the main concern of the modeling work should be in the tails of probability distributions. Yet, when using parametric distributions, the tails are very sensitive to distribution assumptions and to distribution parameters. In the case of discrete distributions without assumption of the form of the distribution, the assessment of tails is still more difficult. These problems are common to all probabilistic approaches, and there have been innumerable attempts

to overcome these problems by fuzzy set theory, rule-based systems, and many other approaches. However, the probabilistic approach (i.e., in risk analysis) appears to become increasingly accepted in practice by administrative bodies and policy makers, and there is a growing demand for efficient techniques for handling probabilistic information.

The belief network approach has, in different versions, been adopted in many fields [3]. Future shows whether the same will occur in the natural resource and environmental sector. There are strong reasons for anticipating that there will be further studies using belief networks, which may include the possibility of performing two-directional, probabilistic computation on-line with reasonable effort and accuracy, compatibility with Bayesian decision analysis and expected utility theory together with compatibility with deterministic management models, and of performing calculations from causes to effects and vice versa [24].

Acknowledgements

This study has been funded and supervised partly by UNU/WIDER under the special Finnish Project Fund, which is supported by the Ministry of Foreign Affairs of Finland, and partly by the International Institute for Applied Systems Analysis, Laxenburg, Austria. Many of the basic ideas have been influenced by Sakari Kuikka from Finnish Game and Fisheries Research Institute. I am grateful to the supporting criticism of the colleagues at the IIASA Water Resources Project, especially to László Somlyódy, David Yates, Kenneth Strzepek and Ilya Masliev. I also want to thank my colleagues at the Helsinki University of Technology and especially Petri Kylmälä who gave valuable comments on the concept of link strength parameter and Pertti Vakkilainen for motivation and support. Also I want to thank the two anonymous referees for their constructive comments.

References

[1] ADB, Environmental Risk Assessment—Dealing with Uncertainty in Environmental Impact Assessment, Asian Development Bank, Manila, 1990.

[2] J.S. Breese, R.P. Goldman and M.P. Wellman, Introduction to special section on knowledge-based construction of probabilistic and decision models, IEEE Trans. Syst. Man Cybernet. 24 (11) (1994) 1577–1579.

[3] D.G. Bobrow, Artificial Intelligence in perspective: a retrospective on fifty volumes of the *Artificial Intelligence Journal*, Artificial Intelligence 59 (1993) 5–20.

[4] R.T. Clemen, Making Hard Decisions, PWS Kent, Boston, MA, 1991.

[5] P. De Jongh, Uncertainty in EIA, in: P. Wathern (Ed.), Environmental Impact Assessment: Theory and Practice, Routledge, London, 1988, pp. 62–84.

[6] J. Gordon and E.H. Shortliffe, A method of managing evidential reasoning in a hierarchical hypothesis space, Artificial Intelligence 26 (1985) 323–357.

[7] E.J. Horwitz, J.S. Breese and M. Henrion, Decision theory in expert systems and artificial intelligence, Internat. J. Approximate Reasoning 2 (1988) 247–302.

[8] R.A. Howard, The foundations of Decision Analysis, IEEE Trans. Syst. Sci. Cybernet. 4 (1968) 211–219.

[9] C. Howson and P. Urbach, Bayesian reasoning in science, Nature 350 (1991) 371–374.

[10] H. Koivusalo, O. Varis and L. Somlyódy, Water Quality of Nitra River, Slovakia—Analysis of Organic Material Pollution, WP-92-084, International Institute for Applied Systems Analysis, Laxenburg, 1992.

[11] G.A. Korn and T.M. Korn, Mathematical Handbook for Scientists and Engineers, McGraw-Hill, New York, 1968.
[12] S. Kuikka and O. Varis, Use of Bayesian Influence Diagram in Fisheries Management—the Baltic Salmon Case, C.M. D:5, International Council for the Exploration of the Sea, 1992.
[13] S. Kuikka and O. Varis, Uncertainties of climatic change impacts in Finnish watersheds: a Bayesian network analysis of expert knowledge, Boreal Environm. Res. 2 (1997) 109–128.
[14] I. Masliev and L. Somlyódy, Uncertainty Analysis and Parameter Estimation for a Class of River Dissolved Oxygen Models, WP-94-9, International Institute for Applied Systems Analysis, Laxenburg, 1994.
[15] M.G. Morgan and M. Henrion, Uncertainty, A Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis, Cambridge University Press, Cambridge, MA, 1990.
[16] D.W. North, A tutorial introduction to Decision Theory, IEEE Trans. Syst. Sci. Cybernet 4 (1968).
[17] J. Pearl, Fusion, propagation, and structuring in belief networks, Artificial Intelligence 29 (1986) 241–288.
[18] J. Pearl, On evidential reasoning in a hierarchy of hypotheses, Artificial Intelligence 28 (1986) 9–15.
[19] J. Pearl, Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference, Morgan Kaufmann, San Mateo, CA, 1988.
[20] H. Raiffa, Decision Analysis, Addison-Wesley, Reading, MA, 1968.
[21] S. Russell and P. Norvig, Artificial Intelligence: A Modern Approach, Prentice-Hall, Englewood Cliffs, NJ, 1995.
[22] R.D. Shachter, Evaluation of influence diagrams, Oper. Res. 34 (1986) 871–882.
[23] R.D. Shachter, Probabilistic inference and influence diagrams, Oper. Res. 36 (1988) 589–604.
[24] R.D. Shachter and D.E. Heckerman, Thinking backward for knowledge acquisition, AI Magazine 8 (1987) 55–61.
[25] G. Shafer, Decision making, in: G. Shafer and J. Pearl (Eds.), Readings in Uncertain Reasoning, Morgan Kaufmann, San Mateo, CA, 1990, pp. 61–67.
[26] G. Shafer and J. Pearl (Eds.), Readings in Uncertain Reasoning, Morgan Kaufmann, San Mateo, CA, 1990.
[27] P.P. Shenoy and G. Shafer, Propagating belief functions with local computations, IEEE Expert (Fall 1986) 43–52.
[28] L. Somlyódy, Quo vadis water quality management in Central and Eastern Europe?, Water Sci. Technol. 30 (1994) 1–14.
[29] L. Somlyódy, I. Masliev, P. Petrovic and M. Kularathna, Water Quality Management in the Nitra River Basin, WP-94-40, International Institute for Applied Systems Analysis, Laxenburg, 1994.
[30] P. Szolovits and S.G. Pauker, Categorial and probabilistic reasoning in medicine revisited, Artificial Intelligence 59 (1993) 167–180.
[31] A. Taskinen, O. Varis, H. Sirviö, J. Mutanen and P. Vakkilainen, Probabilistic uncertainty assessment of phosphorus balance calculations in a watershed, Ecol. Modelling 74 (1994) 125–135.
[32] R.V. Thomann, Systems analysis in water quality management—a 25 year retrospective, in: M.B. Beck (Ed.), Systems Analysis in Water Quality Management Pergamon, Oxford, 1987, pp. 1–14.
[33] O. Varis, Belief networks for modeling and assessment of environmental change, Environmetrics 6 (1995) 439–444.
[34] O. Varis, A Belief Network Approach to Optimization and Parameter Estimation in Resource and Environmental Management Models, WP-95-11, International Institute for Applied Systems Analysis, Laxenburg, 1995.
[35] O. Varis, Interconnections on water, food, poverty, and global urbanization: a qualitative analysis on driving forces, impacts, and policy tools, in: Proceedings International Conference on Large Scale Water Resources Development in Developing Countries: New Dimensions of Prospects and Problems, Kathmandu, Nepal, October 20–23, 1997, Graphlink, Kathmandu, 1997.
[36] O. Varis, Bayesian decision analysis for environmental and resource management, Environmental Modeling and Software 12 (1997) 177–185.
[37] O. Varis, J. Kettunen and H. Sirviö, Bayesian influence diagrams in complex environmental management including observational design, Comput. Statist. Data Anal. 9 (1990) 77–91.
[38] O. Varis and S. Kuikka, Analysis of Sardine Fisheries Management on Lake Kariba, Zimbabwe and Zambia—Structuring and Analysis of a Bayesian Influence Diagram Model, WP-90-48, International Institute for Applied Systems Analysis, Laxenburg, 1990.

[39] O. Varis and S. Kuikka, Joint use of multiple environmental assessment models by a Bayesian meta-model: the Baltic salmon case, Ecol. Modelling 102 (1997) 341–351.
[40] O. Varis and S. Kuikka, BeNe-EIA: a Bayesian approach to expert judgment elicitation with case studies on climatic change impacts on surface waters, Climatic Change 37 (1997) 539–563.
[41] O. Varis, B. Kløve and J. Kettunen, Evaluation of a real-time forecasting system for river water quality—a trade-off between risk attitudes, costs and uncertainty, Environm. Monit. Assessment 28 (1993) 201–213.
[42] O. Varis, S. Kuikka and J. Kettunen, Belief Networks in Fish Stock Assessment—the Baltic Salmon Case, C.M. D:13, International Council for Exploration of the Sea, 1993.
[43] O. Varis, S. Kuikka and A. Taskinen, Modeling for water quality decisions: uncertainty and subjectivity in information, in objectives, and in model structure, Ecol. Modelling 74 (1994) 91–101.
[44] A. Wald, Statistical Decision Functions, Wiley, New York, 1950.
[45] WCED, Our Common Future: Report of the World Commission on Environment and Development, Oxford University Press, Oxford, 1987.
[46] F.H. Zeitz III and P.S. Maybeck, An alternate algorithm for discrete-time filtering. IEEE Trans. Aerosp. Electr. Syst. 29 (1993) 1123–1135.